第26卷，第10期
2006年10月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 26, No. 10, pp1842-1845
October, 2006

近红外光谱法结合支持向量机测定天然牛黄粉中人工牛黄的掺入量

马 群[1,2]，郝贵奇[3]，乔延江[1]，张卓勇[3*]，张孝芳[3]

1. 北京中医药大学中药学院，北京 100102
2. 北京同仁堂股份有限公司科学研究所，北京 100011
3. 首都师范大学化学系，北京 100037

摘 要 提出了应用近红外漫反射光谱技术结合支持向量机测定天然牛黄粉中人工牛黄的含量的方法。以傅里叶变换近红外光谱仪(4 000 ～10 000 cm^{-1})为试验仪器，以含有不等量人工牛黄的天然牛黄粉(天然牛黄的质量分数范围为 0%～100%)作为校正样品，对光谱数据进行平滑、求导和小波压缩，结合支持向量机，建立了测定天然牛黄粉中人工牛黄含量的模型。试验结果为：预测相对误差的平方和可达 0.001 35。研究表明：近红外漫反射光谱法结合支持向量机可以测定天然牛黄粉中人工牛黄的掺入量，结果可靠，可用于天然牛黄粉的质量控制。

主题词 近红外漫反射；支持向量机；天然牛黄；人工牛黄
中图分类号：O657.3 **文献标识码**：A **文章编号**：1000-0593(2006)10-1842-04

引 言

天然牛黄粉为天然牛黄加工而成的粉末，为牛科动物中干燥的胆结石，有多种临床功效，但资源稀少，因此其价格昂贵。人工牛黄粉是根据天然牛黄粉的主要成份配制而成，其原料由人工合成，配制简单，价格便宜，两者药用效果相差很大，但价格悬殊，若在天然牛黄粉中掺入一定量的人工牛黄粉，其理化性质很相似，传统的方法很难鉴别[1]。市场上有人用人工牛黄粉加工后冒充天然牛黄粉或在天然牛黄粉制剂中掺入不等量的人工牛黄粉后出售。因此，天然牛黄和人工牛黄的现代科学方法鉴别及天然牛黄中人工牛黄掺入量的检测对于含牛黄中药产品的质量控制具有非常重要的意义。

目前中药牛黄的质量控制方法仍以经验鉴别为主，辅以少量现代理化分析如薄层色谱、紫外光谱、红外光谱、显微鉴别法等[2]。这些方法的样品预处理较繁琐，费事费力，在一定程度上难以进行快速分析。近红外光谱(NIR)技术是近年来发展较快的光谱分析方法之一。其主要优点是无需复杂的样品前处理，且是一种非破坏性的分析技术，因此对于各种样品的快速检测，特别是生产过程的质量控制，具有特别重要的实际应用意义。近红外光谱的吸收谱带主要是C—H，N—H，O—H 等基团的倍频和合频的吸收，它的谱峰重叠严重，数据的处理和解释非常困难。在早期，由于受到技术水平和实验条件限制，近红外光谱的应用非常有限。近年来，随着计算机和化学计量学技术的发展，近红外光谱分析技术已成功应用于农业[3]、食品工业[4]、石油[5]及制药[6]等领域，尤其适用于在线分析[7]。因其分析速度快、无需前处理、非破坏性及多组分同时定量分析等优势而得到广泛的应用。

支持向量机(support vector machine，简称 SVM)是 Vapnik 等根据统计学习理论提出的一种建立在结构风险最小化原则的基础上，专门研究小样本情况下统计估计和预测的问题，探索在有限样本的情况下如何得到最优解的通用学习方法，体现了兼顾经验风险和置信范围的一种折衷的思想，能较好地解决小样本、非线性和高维数等实际问题。本文利用近红外漫反射光谱技术，运用支持向量回归算法建立数学模型，建立了天然牛黄粉中人工牛黄掺入量的测定方法。结果证明，该方法简便、可靠，具有较大的实际应用价值。

1 支持向量回归(SVR)的原理

支持向量机是 Vapnik 等首先提出并应用的一种新型学习算法[8]。最初，该算法是用来解决模式识别中的二分类问题，在引入 Vapnik 提出的 ε 不敏感损失函数后，支持向量机也可用来解决非线性的回归问题。对于回归建模问题，传统的化学计量学算法在拟合训练样本时，将有限样本数据中的

收稿日期：2006-01-12，**修订日期**：2006-07-13
基金项目：北京市科技发展项目(H040230130710)资助
作者简介：马 群，女，1971年生，北京同仁堂股份有限公司科学研究所高级工程师 * 通讯联系人

误差也拟合进数学模型了，而支持向量回归采用“ε 不敏感函数”，即对于用 $f(x)$ 拟合目标值 y 时，$f(x)=w^{\mathrm{T}}x_i+b$，目标值 y_i 拟合在 $|y_i-w^{\mathrm{T}}x_i-b|\leqslant\varepsilon$ 时，认为进一步拟合是无意义的。这样拟合得到的不是唯一解，而是一组无限多个解。这一求解策略使过拟合受到限制，显著提高了数学模型的预报能力。SVR 方法是在一定约束条件下，以 $\|\omega^2\|$ 取最小的标准为适应样本集的非线性。SVR 通过非线性映射将数据映射到高维的特征空间中，在其中进行线性回归。通过运用一个非敏感性损耗函数，非线性 SVR 的解可通过下面方程[9]求出

$$\max_{a,\,a^*}W(a,\,a^*)=\max_{a,\,a^*}\left\{\begin{array}{l}\sum_{i=1}^{l}a_i^*(y_i-\varepsilon)-a_i(y_i+\varepsilon)\\-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(a_i^*-a_i)(a_j^*-a_j)K(x_i,\,x_j)\end{array}\right\}\tag{1}$$

其约束条件为

$$0\leqslant a_i\leqslant C,\ i=1,\ \cdots,l$$
$$0\leqslant a_i^*\leqslant C,\ i=1,\ \cdots,l$$
$$\sum_{i=1}^{l}(a_i^*-a_i)=0$$

由此可得拉格朗日待定系数 a_i 和 a_i^*，回归函数 $f(x)$ 为

$$f(x)=\sum_{\mathrm{SVs}}(\bar{a}_i-\bar{a}_i^*)K(x_i,\,x)\tag{2}$$

2　实验部分

2.1　仪器与药品

Antaris 傅里叶变换近红外光谱仪(美国 Thermo Nicolet 公司)配有 InGaAs 监测器、积分球漫反射采样系统、Result 操作软件和 TQ Analyst 光谱分析软件。HB-330 型电子天平(日本岛津公司)。中药材天然牛黄和人工牛黄均由北京同仁堂股份有限公司科学研究所提供。

2.2　实验方法

将天然牛黄与人工牛黄分别粉碎后过 80 目筛，然后准确称取天然牛黄粉和人工牛黄粉，配制成不同比例的样品，每份混合均匀。这样共得样品 34 份，天然牛黄质量分数范围为 0%～～100%。利用近红外的旋转器使样品池匀速转动，以积分球漫反射法采集样品的近红外光谱，扫描范围为 4 000 ～10 000 $\mathrm{cm^{-1}}$，分辨率为 8 $\mathrm{cm^{-1}}$，扫描次数为 64。为保证其有代表性，每个样品做 4 次平行实验，求平均光谱。

2.3　数据处理

纯天然牛黄和纯人工牛黄的近红外光谱如图 1 所示。从图上可以看出，有用的光谱信息主要集中在 4 000～7 500 $\mathrm{cm^{-1}}$波段。为从光谱量测数据中充分提取有效信息，减小运算量，首先对该波段光谱采用 Savitzky-Golay 方法进行 2 次多项式 5 点平滑，然后进行一阶和二阶求导，以消除背景和基线的影响。使用 Matlab(Mathworks，Inc.，USA)的内部函数 appcoef 进行一维小波变换，对光谱进行压缩。最后，对压缩后的光谱数据进行归一化。

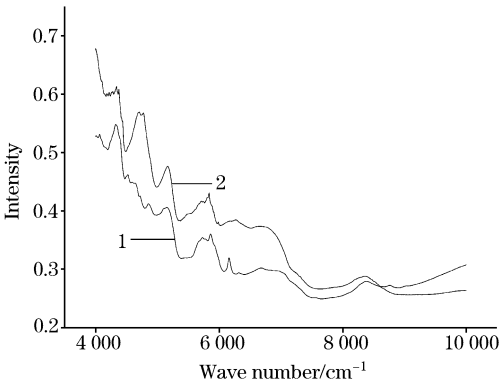

Fig. 1　NIR of bezoar powder and artificial bezoar powder

1: *Bezoar powder*; 2: *Artificial bezoar powder*

3　结果与讨论

本文选择径向基函数作为核函数。用留一法对模型进行验证，所谓留一法就是留一个样本作为预测样本，其他样本作为训练样本，重复此过程，直到每个样本都作为预测样本 1 次，作为预测样本 $n-1$ 次。用平均相对误差平方和来评价模型。

我们在建模时，不把天然牛黄质量分数为 0%和 100%的样品，即纯人工牛黄和纯天然牛黄作为预测样本。

3.1　参数选择

确定了核函数后，所要设定的 3 个参数是惩罚系数 C、不敏感函数 ε 和径向基系数 γ。其中惩罚参数 C 控制着经验风险，一般都取一个较大的数用来降低误差，以取得对训练样本较好的拟合。参数 ε 控制着误差的边界，而参数 γ 则控制着 SVM 对输入量变化的敏感程度。虽然有些启发式的算法可以获取这些参数值，但都不是最理想的，需要通过试验来确定。

表 1 给出了在固定参数 C 和 γ 时，不同 ε 对应的预测结果。随着 ε 的减小，预测评价相对误差也逐渐减小，当 ε 减小到 0.001 后，预测误差变化不再明显。表 2 给出了固定参数 C 和 ε，调整 γ 对应的预测结果。过大的 γ 使 SVM 反应迟钝，而过小的 γ 对输入过于敏感，致使样本预测结果也不好。在确定了 ε 和 γ 之后，改变 C 的值，当 C 在较大的范围内变动时，

Table 1　Influence of paramter ε on the result of support vector regression

ε	相对误差平方和
1.5×10^{-1}	6.89
5.0×10^{-2}	8.11×10^{-1}
1.0×10^{-2}	4.18×10^{-1}
5.0×10^{-3}	1.64×10^{-1}
1.0×10^{-3}	5.78×10^{-3}
5.0×10^{-4}	5.60×10^{-3}
1.0×10^{-4}	5.64×10^{-3}
5.0×10^{-5}	5.68×10^{-3}
1.0×10^{-5}	5.71×10^{-3}

注：$C=10$，$\gamma=1$

预测结果没有变化，这也表明 ε 和 γ 的选取是合适的。但是，最好在此基础上适当减小 C 的值，避免过大的 C 值引起经验误差，导致泛化能力下降。

Table 2 Influence of γ on the result of support vector regression

γ	相对误差平方和
1×10^{-3}	5.05×10^{-2}
5×10^{-3}	5.02×10^{-3}
1×10^{-2}	1.35×10^{-3}
5×10^{-2}	2.11×10^{-3}
1×10^{-1}	2.78×10^{-3}
3×10^{-1}	1.72×10^{-3}
5×10^{-1}	3.09×10^{-3}
1	5.64×10^{-3}
2	1.06×10^{-1}
5	2.71

注：$C=10$，$\varepsilon=0.0001$

3.2 光谱预处理方法对预测结果的影响

对光谱进行求导可以消除基线和背景的干扰，但同时会放大噪声，尤其是二阶导数光谱。此外，小波压缩的次数也会对预测结果产生一定的影响。本文考察了一阶导数、二阶导数和小波压缩次数对预测结果的影响，结果见表 3。不同的数据预处理方法对结果的影响较大。采用二阶导数和 4 次小波压缩预测的误差最大，一阶导数和 5 次小波压缩的误差最小。本研究采用一阶导数和 5 次小波压缩相组合的数据预处理方法。最终预测结果见表 4，样品的预测误差在 −0.712%～0.578%之间。

Table 3 Influence of methods of data process on the result of support vector regression

数据处理方法		相对误差平方煌
一阶导数	4 次小波压缩	0.150
	5 次小波压缩	0.00135
二阶导数	4 次小波压缩	3.87
	5 次小波压缩	0.461

4 结 论

本文将近红外光谱与支持向量机相结合，建立了预测天然牛黄粉中人工牛黄掺入量的模型。样品粉碎后无须进行复杂处理就可以用近红外光谱仪进行测定。对 NIR 光谱进行平滑、求导和小波压缩后，用支持向量回归算法建立模型。用留一法进行检验，样品的预测误差在 −0.712%～0.578%之间。该方法可用于天然牛黄的质量控制，并有望应用于其他中药材的质量控制。

致谢：本研究采用的支持向量机软件 ChemSVM 1.0 由上海大学计算机化学研究室陆文聪教授提供，特此致谢。

Table 4 Prediction results of model built by support vector regression

样品号	真值/%	预测值%	误差/%	相对误差/%
1	98.04	98.08	0.0387	0.0395
2	96.15	96.20	0.0502	0.0522
3	94.34	94.32	−0.0180	−0.0191
4	92.59	92.22	−0.365	−0.395
5	90.91	90.48	−0.429	−0.472
6	89.29	88.94	−0.347	−0.389
7	86.21	86.41	0.199	0.231
8	83.33	83.57	0.245	0.293
9	80.65	80.69	0.0439	0.0544
10	78.13	78.12	−0.0144	−0.0184
11	75.76	75.73	−0.0277	−0.0366
12	73.53	73.52	−0.0105	−0.0143
13	71.43	70.88	−0.550	−0.770
14	69.44	68.73	−0.712	−1.03
15	67.57	67.30	−0.265	−0.392
16	65.79	66.25	0.464	0.705
17	64.10	64.35	0.248	0.387
18	62.50	62.70	0.202	0.324
19	59.52	59.34	−0.179	−0.300
20	55.56	55.41	−0.146	−0.262
21	51.02	50.91	−0.113	−0.222
22	48.08	48.21	0.130	0.271
23	44.64	45.22	0.578	1.30
24	42.37	42.91	0.541	1.28
25	40.00	40.35	0.345	0.863
26	36.36	36.80	0.439	1.21
27	28.57	28.31	−0.264	−0.923
28	25.00	24.60	−0.396	−1.58
29	20.00	20.09	0.0906	0.453
30	16.67	16.69	0.0204	0.122
31	12.50	12.51	0.00610	0.0488
32	9.090	9.005	−0.0851	−0.936

参　考　文　献

[1] LIN Pei-ying, WANG Hong, DONG Xin, et al(林培英, 王　洪, 董　昕, 等). Journal of the Chinese Medicinal Materials(中药材), 2005, 28(3): 177.

[2] ZHAO Bao-gui, LIU Hong, DUN Jin-ye(赵宝贵, 刘　红, 顿金叶). Shaanxi Journal of Traditional Chinese Medicine(陕西中医), 2000, 29(1): 38.

[3] LU Bao-hong, ZHANG Jun, ZHANG Yi-rong, et al(卢宝红, 张　俊, 张义荣, 等). Journal of the Chinese Cereals and Oils Association(中国粮油学报), 2005, 20(4): 44.

[4] CHEN Bin, YU Li-yan, LU Dao-li, et al(陈　斌, 于丽燕, 陆道礼, 等). China Condiment(中国调味品), 2003, 9: 40.

[5] GAO Jun, XU Yong-ye, YAO Cheng(高　俊, 徐永业, 姚　成). Journal of Nanjing University of Technology(南京工业大学学报), 2005, 27(3): 51.

[6] REN Rui-xue, TANG Zhen, LIU Fu-qiang, et al(任瑞雪, 汤　真, 刘福强, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2001, 21(4): 521.

[7] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立, 袁洪福, 陆婉珍). Modern Scientific Instruments(现代科学仪器), 2004, (2): 3.

[8] ZHANG Lu-da, SU Shi-guang, WANG Lai-sheng, et al(张录达, 苏时光, 王来生, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2005, 25(1): 33.

[9] LU Wen-cong, CHEN Nian-yi, YE Chen-zhou, et al(陆文聪, 陈念贻, 叶晨洲, 等). Computers and Applied Chemistry(计算机与应用化学), 2002, 19(6): 701.

Determination of the Artificial Bezoar Powder in Bezoar Powder by Near-Infrared Spectrometry and Support Vector Machine

MA Qun[1, 2], HAO Gui-qi[3], QIAO Yan-jiang[1], ZHANG Zhuo-yong[3*], ZHANG Xiao-fang[3]

1. Beijing University of Chinese Medicinal and Pharmaceutical Sciences, Beijing　100102, China

2. Research Institute, Beijing Tongrentang Co., Ltd, Beijing　100011, China

3. Department of Chemistry, Capital Normal University, Beijing　100037, China

Abstract　A method for determining the artificial bezoar powder in bezoar powder using near-infrared(NIR) diffuse reflectance spectrometry was proposed in the present paper. The method was based on support vector machine (SVM). The calibration set was set up by adding unequal artificial bezoar powder to the bezoar powder (content range: 0%-100%) and collecting the NIR spectrum of the samples in the wave number range of 4 000-10 000 cm^{-1}. The processing algorithm was wavelet transform with first and second derivatives. A mathematical model with support vector machine was established. The model was checked with leave one method. The sum of the square of the relative prediction error was 0.001 35. This method is reliable and can be used to control the quality of bezoar powder.

Keywords　Near-infrared diffuse reflectance; Support Vector Machine; Bezoar powder; Artificial bezoar powder

(Received Jan. 12, 2006; accepted Jul. 13, 2006)

* Corresponding author